DOCUMENT RESUME

ED 258 493 HE 018 401

AUTHOR Yost, Michael
TITLE Predicting Attrition in Admissions Using a Discriminant Model. SAIR Conference Paper.
PUB DATE Oct 84
NOTE 15p.; Paper presented at the Annual Conference of the Southern Association for Institutional Research (Little Rock, AR, October 24-26, 1984). For related document see HE 018 402.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Admission; *College Applicants; *College Attendance; *College Freshmen; *Discriminant Analysis; Higher Education; *Mathematical Models; *Prediction; *Student Attrition
IDENTIFIERS *No Shows; SAIR Conference; Trinity University TX

ABSTRACT

A mathematical model designed to identify admitted college applicants who pay a tuition deposit but fail to attend the college was tested. A two-group discriminant model was built using 486 Trinity University students who paid deposits and enrolled as freshmen during the 1983-1984 academic year, and 72 persons who paid tuition deposits but did not enroll as freshmen. The variables included in the discriminant model were: verbal and mathematics Scholastic Aptitude Test scores, high school rank, high school grade point average, sex, amount of financial aid provided, and place of residence (in-state or out-of-state). For the 1983-1984 year, a significant difference was found between the two groups. The discriminant weights from this model were then applied to persons who paid deposits for entrance into the 1984-1985 freshman class. In this academic year, the model tested 610 students who paid deposits and enrolled and 79 persons who paid deposits but did not enroll. The second statistical analysis validated the previous year findings of a significant difference between the two groups. The accuracy of prediction was 75 percent for the 1983-1984 group and 68 percent for the 1984-1985 group. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED258493

Predicting Attrition in Admissions Using a Discriminant Model

by

Michael Yost, Ph.D.

Trinity University

Paper presented at the annual conference of the Southern Association for Institutional Research

Little Rock, Arkansas

October 1984

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

SAIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SOUTHERN ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the 1984 Annual Conference of the Southern Association for Institutional Research held in Little Rock, Arkansas, October 24-26, 1984. It was reviewed by the SAIR Publications Committee and was judged to be of interest and pertinent to others concerned with the research in higher education. This paper has therefore been selected to be included in the ERIC collection of Conference Papers.

Richard D. Howard
President, SAIR

Predicting Attrition in Admissions Using a Discriminant Model

Abstract

With the current emphasis placed on recruiting students, many administrators are interested in the early identification of prospective students who pay tuition deposits and either request a refund of the deposit or do not enroll, even though they have paid a deposit. The purpose of this research was to determine whether a statistical model could be developed which would identify those individuals who pay tuition deposits at an institution and either request it back or do not enroll. Further, if a statistical model could be developed for one year, how well would this model fit a second year. A two-group discriminant model was built using those persons who paid deposits and enrolled as one group, and those persons who paid deposits and either requested it back or did not enroll as the other group. The variables included in the discriminant model were: SAT Verbal, SAT Math, high school rank, high school GPA, sex (dichotomous), amount of financial aid offered, and location of home (in-state, out-of-state). The discriminant model was run on those persons who paid deposits for entrance into the 1983-84 freshman class. The probability level for the model was less than .01 and there was an overall prediction accuracy of 75 percent. The discriminant weights from this model were then applied to those persons who paid deposits for entrance into the 1984-85 freshman class. There was a 68 percent overall accuracy of prediction rate with the 1983-84 freshman class using the weights obtained with the 1982-83 class. These results indicate that it is possible to statistically discriminate between those persons who pay deposits and do enroll and those who do not enroll. The results also indicate that the model is applicable to applicants from each of two years.

Predicting Attrition in Admissions Using a Discriminant Model

Introduction

In the last few years, the number of people of college entrance age has tended to decrease. This decrease in numbers, along with a number of other factors, have led to an increase in the amount of effort being made by colleges and universities to recruit students. As more effort is put into recruiting students, the students themselves are becoming more aware that they are important to the colleges and universities to which they apply for admission. Many students have become keen shoppers for the institutions to which they will apply and eventually attend.

Today's students spend a great deal of time looking for what they consider to be the "right" school for them and, in many instances, for schools which are able to offer them good financial aid packages. As students seek out what they consider to be the best school for them, they tend to gather more information from more schools than in the past, to apply to more schools, to be accepted for admission by more schools, and to pay tuition deposits at more schools.

The increases in the number of applications and tuition deposits received at institutions has caused problems. The inflated number of applications is inconsistent with historical trends at many institutions, and this has made it difficult to predict the size of the freshman class. An even more difficult

situation for institutions is one in which many students pay tuition deposits to more that one institution, even though they will attend only one of the institutions. Many of the students who pay tuition deposits at more than one institution do so to assure themselves of admission at one institution while they wait to hear from the school which is their first choice.

The problem which is created for the colleges and universities which receive tuition deposits from students who pay them to more than one institution is that the institutions are not able to identify these individuals, and it is difficult to tell when to close admissions, and how many accomodations to prepare for incoming freshmen.

The research described in this article is a two-year study in which a mathematical model was developed and validated, and which can be used to identify a significant number of those individuals who pay a tuition deposit to an institution and do not attend.

Problem

Trinity University is a relatively small (approximately 3000 students) private institution located in San Antonio, Texas. In the past few years, our enrollment has tended to remain stable, while the number of applications for admission and the number of

tuition deposits that we received have increased. In each of the last two years, approximately 18 percent of the persons who have paid tuition deposits have fallen into one of three categories. First, there are those students who requested that their tuition deposit be returned and have had it returned. Second, there are those students who requested that their tuition deposit be returned and have not had it returned because they waited past a deadline to make their request. Third, there are those who never requested that their tuition deposit be returned and simply did not appear for advising and registration. The fact that this large a proportion of the students who paid deposits did not enroll has made it difficult to predict the size of the freshman class, has made it difficult to determine when to close admissions, and has made it difficult to estimate housing needs. Therefore, this study was conducted to determine whether there was any way of identifying those students who paid tuition deposits and did not enroll.

Design

A graphic description of the design of this study is contained in Figure 1. In the first half of the study, those students who paid tuition deposits for entrance into the 1983-84 freshman class were the subjects. In this academic year, 520 students paid deposits and enrolled as freshmen, while 91 students paid tuition deposits and did not enroll as freshmen.

Of the 520 and 91 students, complete data were available on 486 (Group B) and 72 (Group A) students, respectively.

* * * * * * * * * * * * * * * * * * *

Insert Figure 1 about here.

* * * * * * * * * * * * * * * * * * *

Eight variables were used in the study. The names of the variables are contained in Figure 1. The SAT verbal score, the SAT quantitative score, the high school grade point average, the month in which an application was received, and the month in which the tuition deposit was recieved are all continuous variables. The meaning and use of the SAT scores and the GPA scores are self-evident and need no explanation. The two scores dealing with the date on which the application and deposits were received does require an explanation. Applications and deposits received between the first of July and the fifteenth of July in 1982 were given a value of 1, while applications and deposits received between the sixteenth of July and the last day of July were given the value of 2. This scheme of assigning numeric values to the first and second half of each month was continued through the opening of the 1983-84 academic year.

The sex, financial aid, and scholarship variables were dichrotomous, and a vector of ones and zeros was developed for each variable. That is, if a person was a male, then his score on this variable was a 1, and otherwise it was a 0. The same approach was used for those students who were offered scholarships or financial aid. Each subject had a value of 1 or 0 in each of the three dichotomous vectors.

In the second half of the study, those students who paid deposits for entrance into the 1984-85 freshman class were the subjects. In this acedemic year, 630 students paid tuition deposits and enrolled as freshmen, while 88 students paid tuition deposits and did not enroll as freshmen. Complete data were available on 610 and 79 of the 630 and 88 students respectively. The second half of the study was a replication of the first half, and was conducted to show the validity of the results.

Analysis and Results

The data were submitted to a discriminant analysis using the Statistical Package for the Social Sciences (SPSS). A complete (not a stepwise) solution was used. A summary of the results of the statistical analysis are contained in Table 1. There was a significant difference between the two groups using the eight variables. In a univariate analysis of variance sense, there was

a significant difference between the two groups on three of the eight variables. Note that the three dichotomous variables were not included in the univariate analysis of variance portion of the table. The hit-and-miss classification data for the discriminant analysis was:

	Predicted to enroll	Predicted to not enroll
Enrolled (B)	364	122
	74.9%	25.1%
Not Enrolled (A)	17	55
	23.5%	76.4%

(Overall accuracy of prediction of 75%)

A set of 'B' weights for the eight variables was obtained from the analysis of the data for those persons who applied for admission to the 1983-84 freshman class. The data for those persons who applied for admission to the 1984-85 freshman class were then prepared for analysis. Complete data were available for 610 of the 629 persons who paid tuition deposits and enrolled, and for 79 of the 88 persons who paid deposits and did not enroll. The 'B' weights obtained from the analysis of the 1983-84 data were then multiplied by the 1984-85 data and a weighted linear composite score was obtained for each of the 689 persons. The hit-miss ratio for the second set of data using the

'B' weights from the first data were:

	Predicted to enroll	Predicted to not enroll
Enrolled (B)	414	196
	67.8%	32.2%
Not Enrolled (A)	25	54
	31.6%	68.3%

(Overall accuracy of prediction of 67%)

* * * * * * * * * * * * * *

Insert Table 1 about here

* * * * * * * * * * * * * *

Conclusions

The admissions problem which prompted this research was that it was difficult to estimate and predict the size of the freshman class from the number of tuition deposits received at our institution. That is, the shrinkage in the number of tuition deposits received from freshmen students made it difficult to predict the size of the freshman class and this, in turn, caused

problems in terms the type and amount of preparation to make for a freshman class of an undetermined size. There was also the possibility that, if those persons who paid tuition deposits and were not going to enroll could be identified early in the year, they could be communicated with and their attrition from the list of expected enrollees could be avoided.

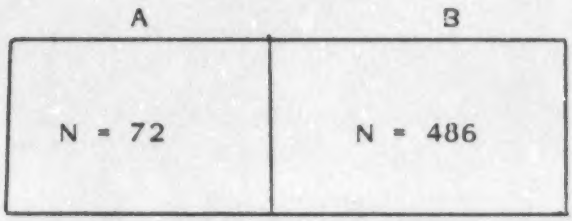
The multivariate hypothesis tested was that there was no significant difference between those students who paid deposits and enrolled (Group B) and those students who paid deposits and did not enroll (Group A). Using eight variables for each student who paid a tuition deposit for admission to the 1983-84 freshman class in a discriminant analysis, the null hypothesis was rejected at the .001 level. This indicated that there was, in fact, a difference between those students who paid deposits and enrolled and those who paid deposits and did not enroll. It then remained to be determined whether this difference was spurious, or whether this difference would exist for students who paid deposits in some other year. The 'B' weights obtained from the discriminant analysis of the 1983-84 data were then applied to the data for the students who paid deposits to enter the 1984-85 freshman class. This statistical approach was an attempt to determine the goodness of fit of the results of the statistical analysis of the 1983-84 data for the 1984-85 data. The second statistical analysis validated the first statistical analysis and indicated that the differences between the two groups of students was consistent for two consecutive years.

The fact that the accuracy of prediction of group membership dropped from 75% with the 1983-84 data to 68% with the 1984-85 data was expected. This reduction was a result of differences in the samples, and is traditionally expected with this type of statistical analysis.

The results of this study are not strong enough to be able to predict the size of an incoming freshman class at some time during an academic year. However, the results are strong enough to indicate that those students who do pay tuition deposits and do not enroll do have something in common, which makes them different from those students who pay tuition deposits and do enroll. The 'B' weights generated in this study can be applied to the data of those students who apply for admission to the 1985-86 freshman class to obtain a weighted linear composite score. Using the weighted linear composite scores and the criteria for predicted group membership obtained in this study, we will be able to identify those students who have the highest probabilities of paying tuition deposits and not enrolling. These students can then be targeted for contact by admissions counselors, financial aid counselors, and other university persons who can work to help these students to enroll and not be lost from our freshman class.

Figure 1

Experimental Design



A = Persons who paid deposits but did not enroll

B = Persons who paid deposits and enrolled

Variables Used in the Discriminant Analysis

1. SAT verbal score
2. SAT quantitative score
3. High school grade point average
4. Sex (dichotomous variable where 1 = male, 0 = female)
5. Month in which the application was received
6. Month in which the deposit was received
7. Scholarship offered (dichotomous variable where 1 = yes, 0 = no)
8. Financial aid offered (dichotomous variable where 1 = yes, 0 = no)

Table 1

Statistical Summary

Means of Continuous Variables

Group	SAT Verbal	SAT Quan.	High School GPA	Applic. Month	Deposit Month
A	534	596	3.52	9.4	14.2
B	552	576	2.96	10.8	15.3

Proportions of Dichotomous Variables

Group	% Male	% Offered Scholarships	% Offered Aid
A	44%	45%	45%
B	48%	65%	46%

Univariate ANOVA for Continuous Variables

Variable	F(1,556df)	Probability
SAT Verbal	2.78	.095
SAT Quan.	3.64	.056
High School GPA	3.66	.056
Application Month	.14	.700
Deposit Month	9.93	.002

Multivariate Test of Significance

Wilks' Lambda	D.F.	Probability
.9577	8	.001